Process Monitoring and Basic Statistics

Richard Metcalf
Idaho National Laboratory

LLNL-INL Safeguards Training Program

July 23, 2009

Idaho National Laboratory

Disclaimer

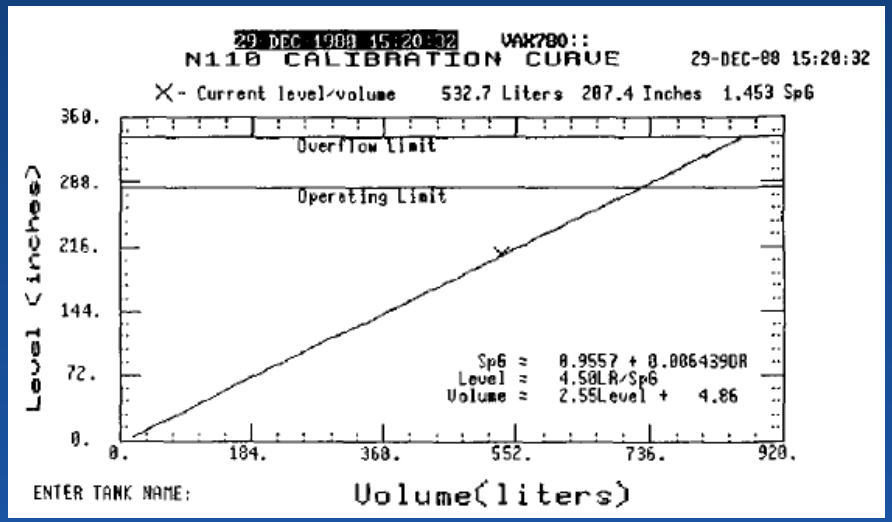
- Process Monitoring is a very large field, and I will not cover everything involved in the application
- Additionally, Statistics is a very precise science, with very specialized vernacular.
  - Pursuant with demystifying the vernacular, common patois will be employed

Overview

- What is Process Monitoring?
  - Where is it used?
  - What does it look like?
- Statistics
  - Basic Statistics Terminology
  - Random and Systematic Error
  - Gaussian (Normal Distribution)
  - Z, Students t, and significance
  - Cautions about Statistics

What is Process Monitoring (PM)?

- Observation of operating characteristics
  - Useful to Operators (is my plant running efficiently?)
  - Useful to Regulators (is the plant running safely?)
  - Useful to Safeguards (is the plant running securely?)

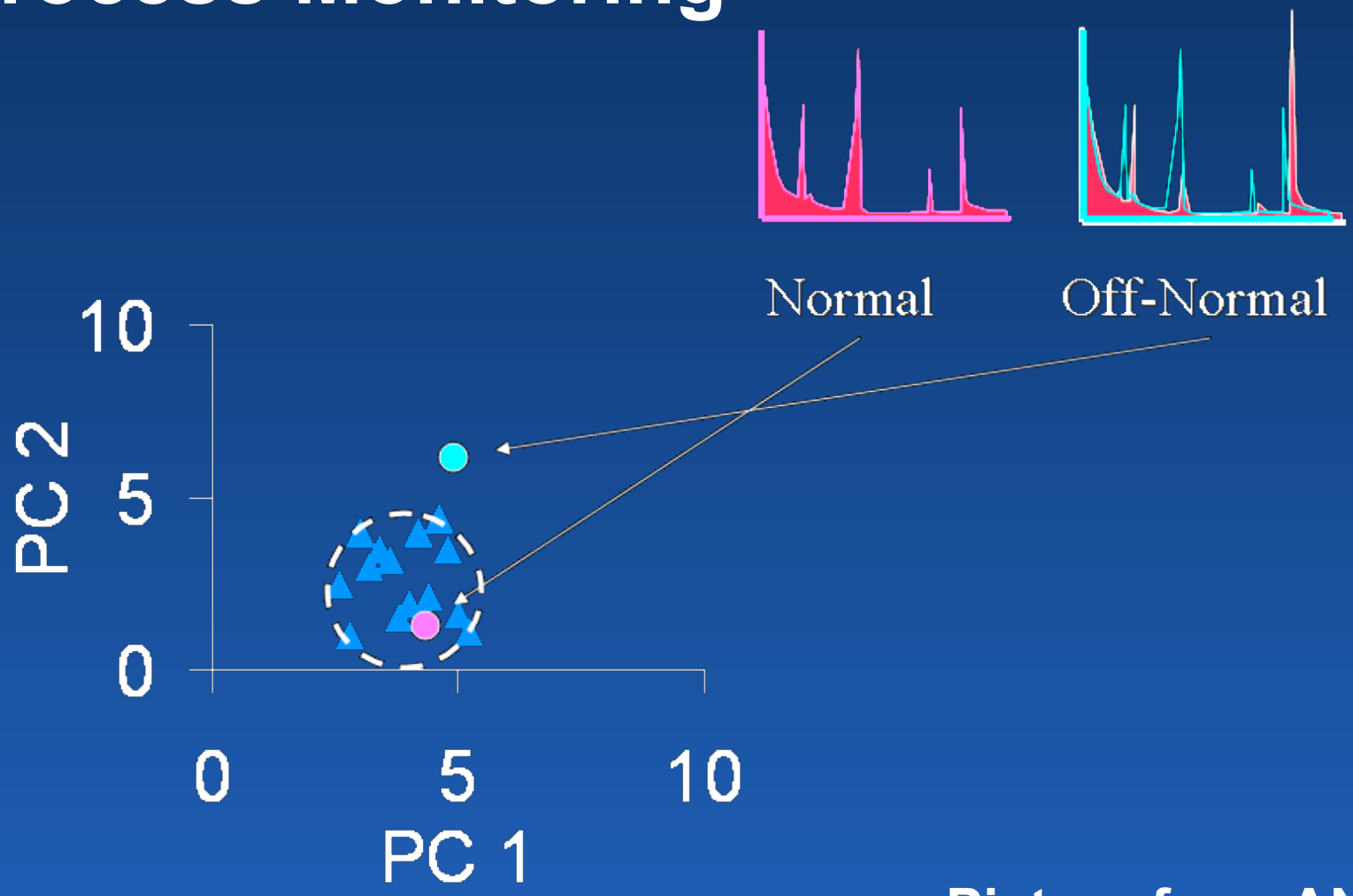


Idaho National Laboratory

Process Monitoring Cont

- Supplement to existing MC&A methods
  - Increase “Safeguardability”
  - Provide Additional Confidence
  - Protect Against Different Threats
  - Potentially allow for larger facilities
  - Reduce Bias Error in DA methods
- Provide an early indication that something off
- Pinpoint areas for greater inspector attention

Process Monitoring

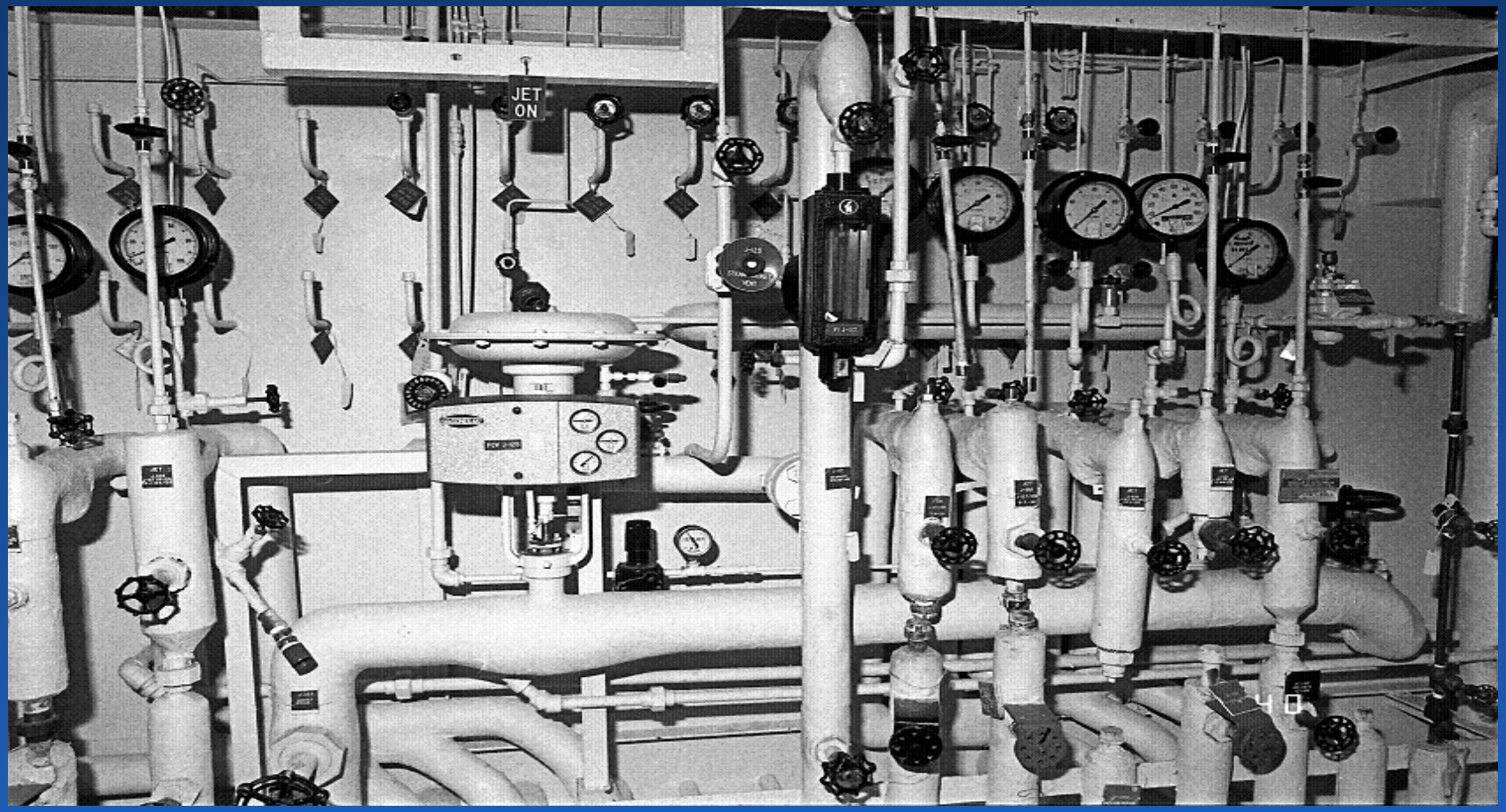


Picture from ANL

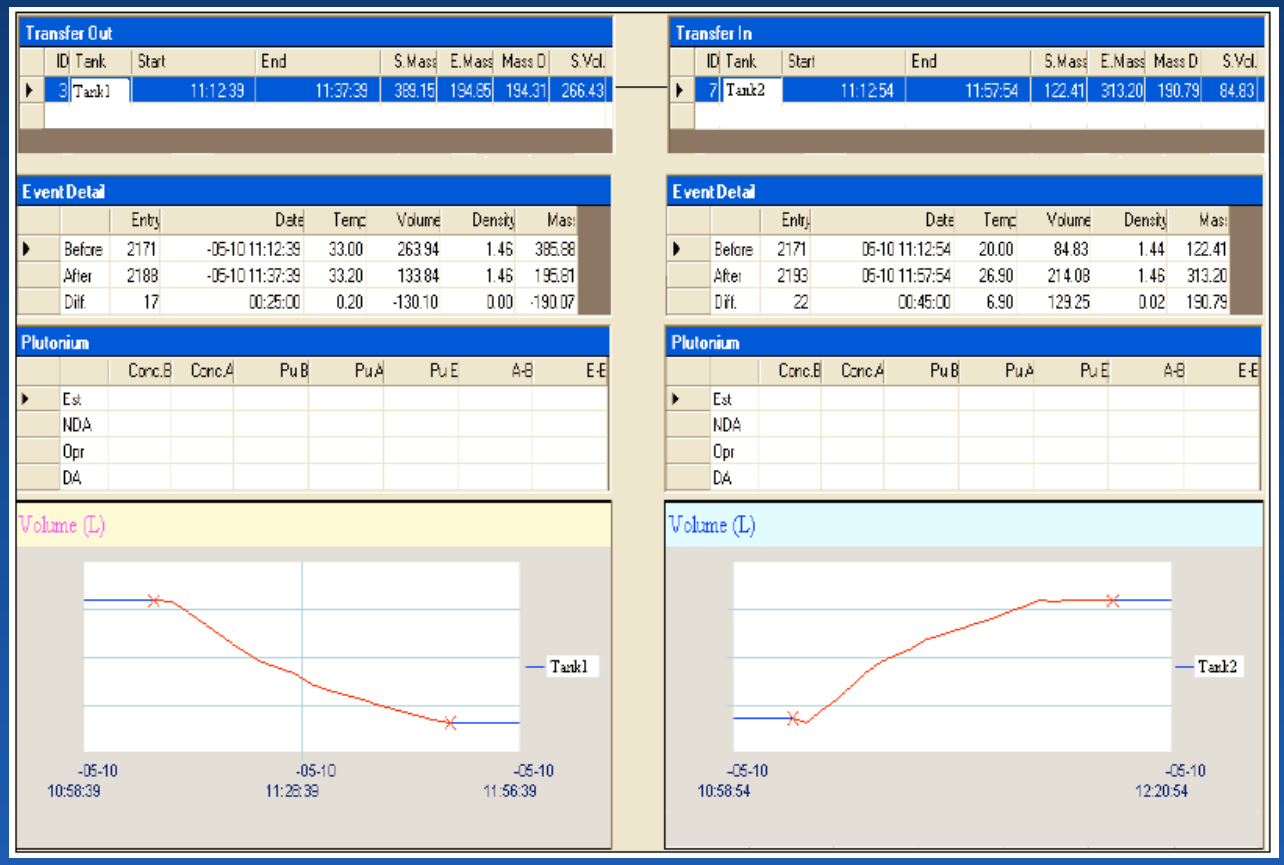
INL Idaho National Laboratory

Process Monitoring: Integrated Equipment Test Facility (ORNL)

Processing Monitoring: Idaho Chemical Processing Plant



Process Monitoring, TAMES



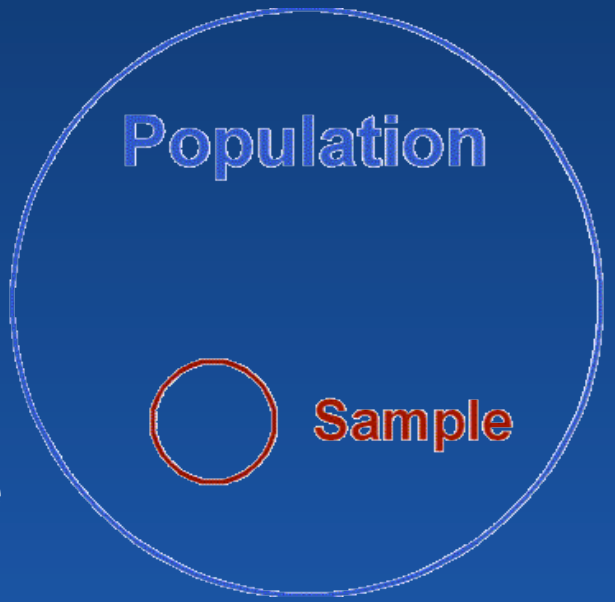
Idaho National Laboratory

Process Monitoring and Statistics

- Large Facilities always require application of statistics
  - How many barrels do I need to open for a 95% confidence that they contain 42g each of Pu?
  - How many DA samples must I take to be sure that the solution has the right concentration?
- Process Monitoring is even more statistics intensive
  - Often correlating tank levels, volume, temperature, pH (less common),

Statistics I: Basic Terminology

- Population: Entire collection of items that is the focus of concern
  - Students throughout our entire lecture series
- Sample: Set of a population that we use to make inferences about the entire population
  - Preferred in most cases because the population is too large or too difficult to measure
  - Students at INL attending the lecture series



Statistics: True Value

- *JARGON WARNING*
- "True" value sets are measurements about a Population
  - *This can be very confusing*
  - How much plutonium is in this tank is a "true" value
    - What we measure is not the "true" value,
    - We infer a confidence about the "true" value
  - Unfortunately, this term can be misused commonly in literature
    - The random error associated with a measurement without an estimate of the bias error can often be mistakenly referred to as a true set

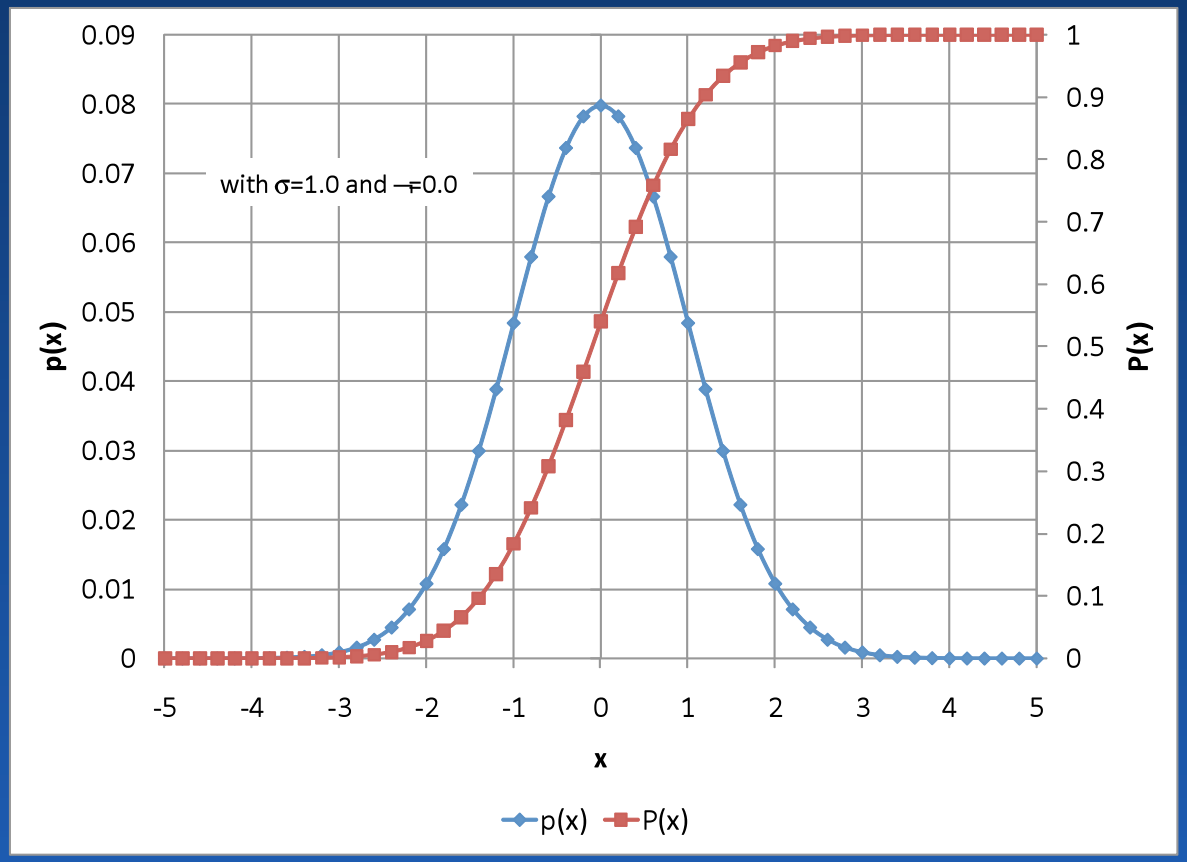
Statistics: Distributions II

- Measurements around a “true” value will always show some error. For this lecture, lets assume that all of our error will be “normal”
  - That is not to say that there are not other distributions of error
  - That is also not to say that we will not see different *types* of error, just that they have the “normal” property
- Normal error is a *Gaussian Distribution*

INL Idaho National Laboratory

Normal Distribution

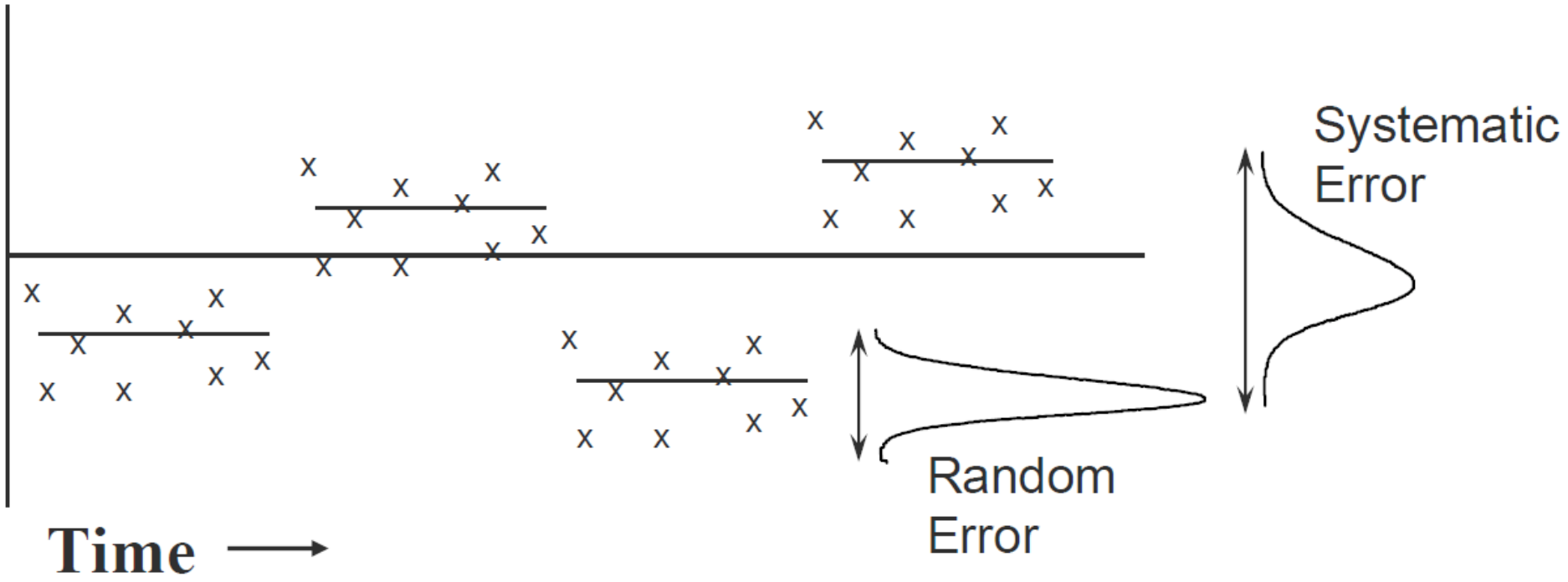
- p(x) is the probability about x that our value will be found
- P(x) is the cumulative probability that x is within the set



INL Idaho National Laboratory

Statistics: Types of Error

- Error in most cases is either systematic (bias) error or random error



Statistics II: True Population Values

- Mean, Mode, and Median: Average, Most Common, Middle Value
- Average Deviation (uncommonly used): the average deviation from the mean value

$$\frac{1}{n}\sum_{i=1}^{n}\left|(x_i - \bar{x})\right|$$

- Standard Deviation (common): Root-Mean-Square of the deviation

Population Standard Deviation

$$\sigma = \sqrt{\sum_{i=1}^{N}\frac{(x_i - \bar{x})^2}{N}} = \sqrt{\sum_{i=1}^{N}\frac{x_i^2}{N} - \mu^2}$$

INL Idaho National Laboratory

Statistics III: Variance

- Variance: square the standard deviation
  - Several interesting properties, such as an additive property
  - The basic “unit” that people tend to use to describe samples
  - Can still be deceptive
    - Outliers can greatly affect a standard deviation (keep in mind that square has a big effect)

Statistics IV: Samples

- **Sample Standard Deviation: measure of the standard deviation of the sample**
  - **The n-1 is related to the *degrees of freedom*,**

Sample Standard Deviation

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n - 1}}$$

  - **As n (sample size) approaches N (population size), this approaches the true standard deviation**

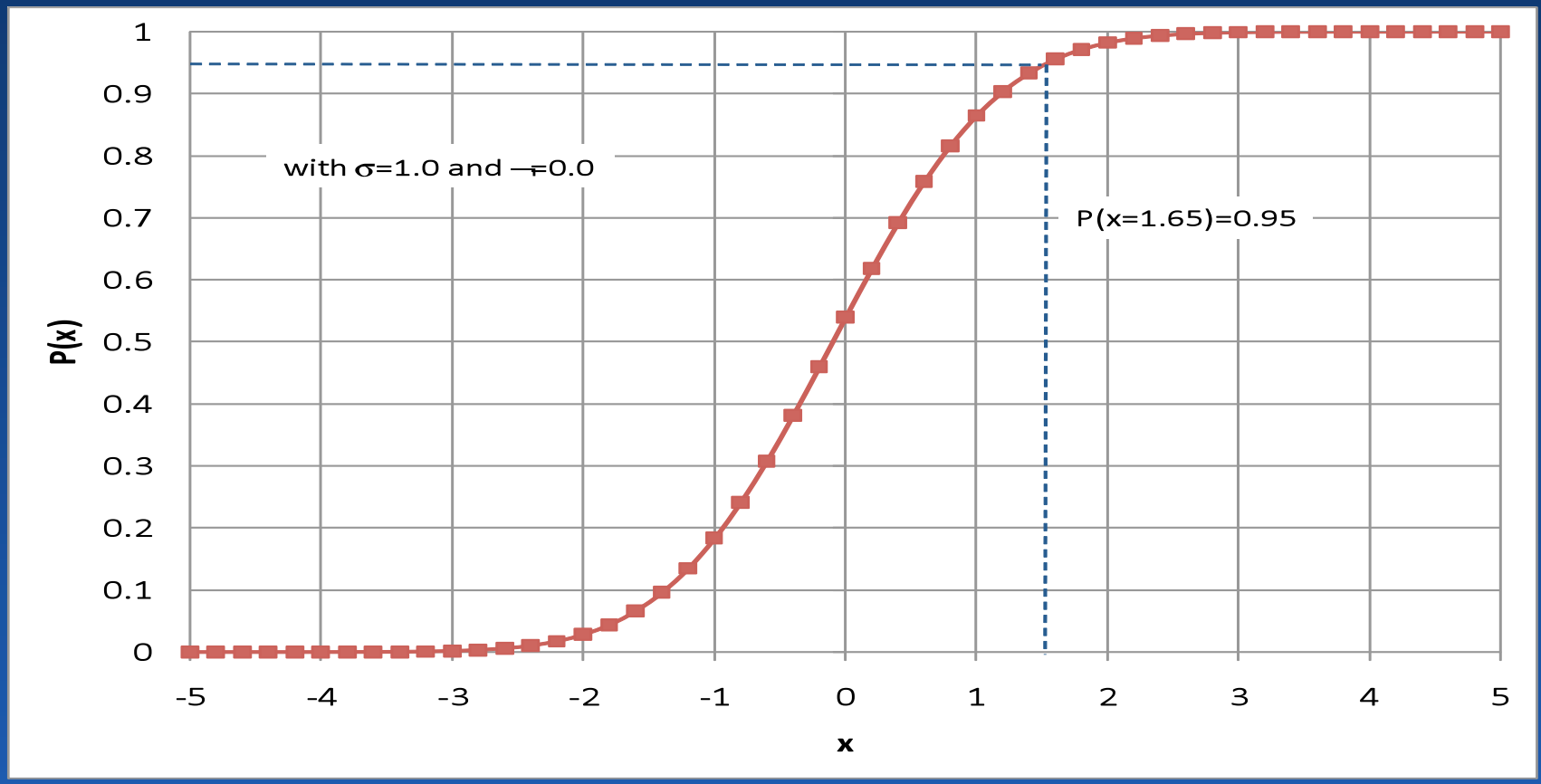
Statistics: Z and Students t Testing

- Z testing: comparison of a sample and population mean to determine if there is a significant difference
  - Must be a test between two values
  - First value is an assumed value called the “null hypothesis”
  - Second value will be a calculated *Z number* or *Z test statistic*
  - Z testing is the most basic statistical test
  - Requires a known population deviation and mean
    - This is somewhat statistically cheating as we never actually have those numbers

Statistics: Students t Test

- Students t test is a test that uses the degrees of freedom based on the number of measurements to make assumptions about the normal distribution in comparing a sample to a believed normal
  - Invented by Guinness in Ireland for better stout
  - Uses some implicit assumptions related to the chi square distribution, normality in the error, and others which are significantly more arcane
  - Works almost exactly like a Z test
  - Recommended as the test of choice for most applications

Statistics: Z Statistics



Z and Students t tests

- Comparison of Population Means
- Determines likelihood of measured means the same or different
- Compared to a threshold, significance value (p value)

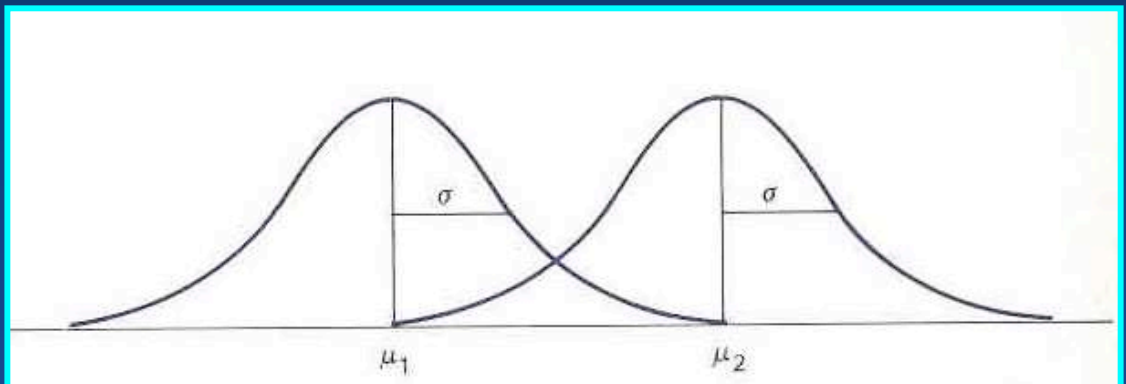


Figure 8.9 Normals with different means, same variance.

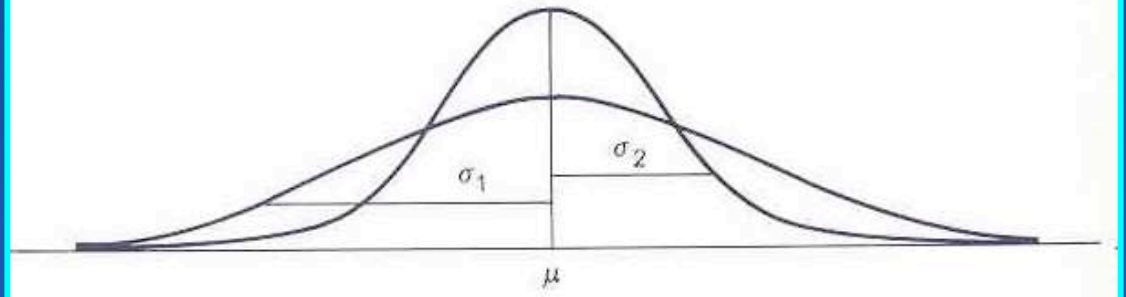


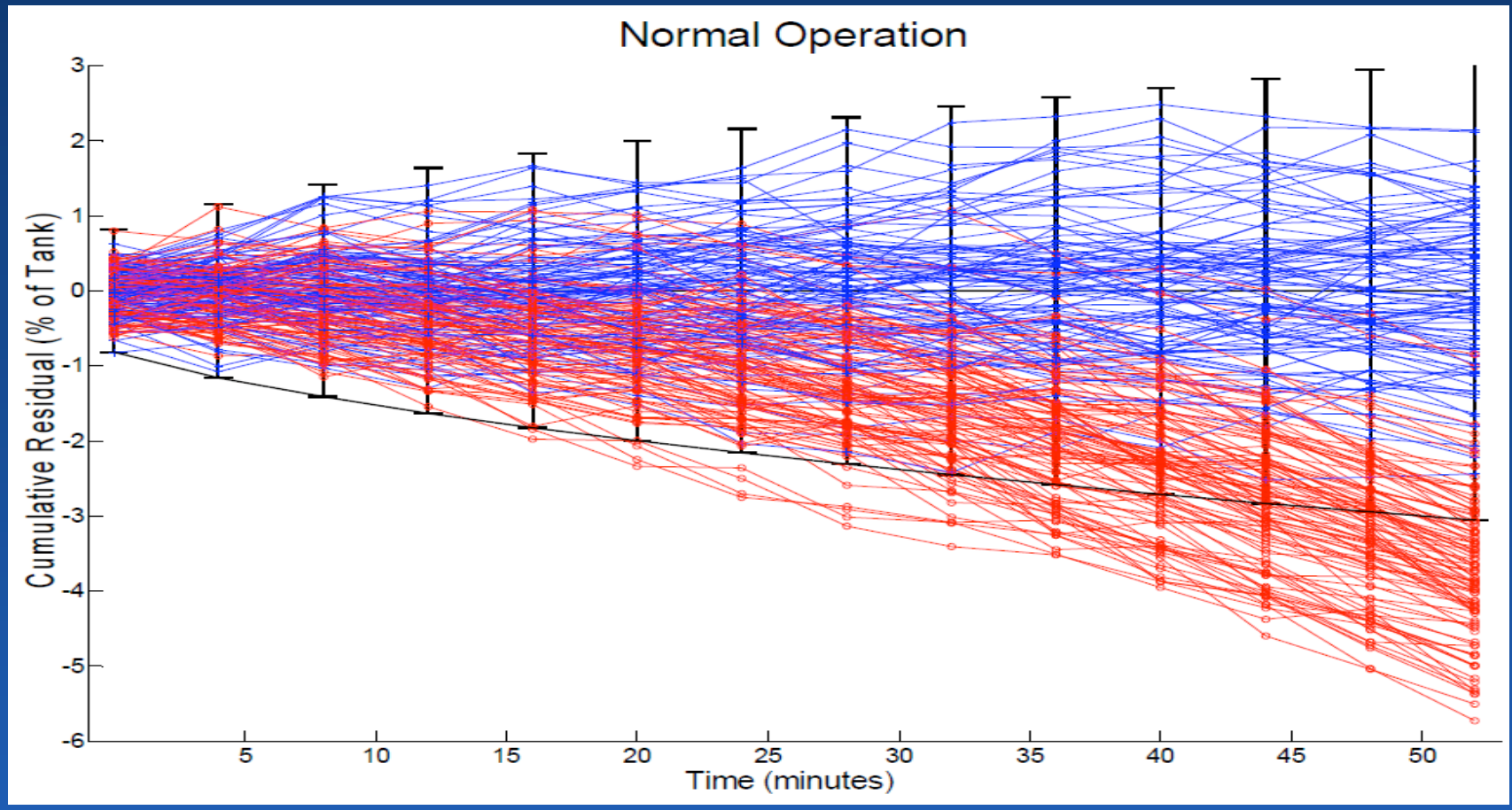
Figure 8.10 Normals with different variances, same mean.

Z, T, P and Confidence Intervals

- Z and t statistics are compared to a P value
- P values are the likelihood that the result was a function of chance
  - P values are small for high confidence
  - P values are a typical reporting mechanism when comparing two groups
- Confidence Intervals are values that we can reasonably expect to contain the true values

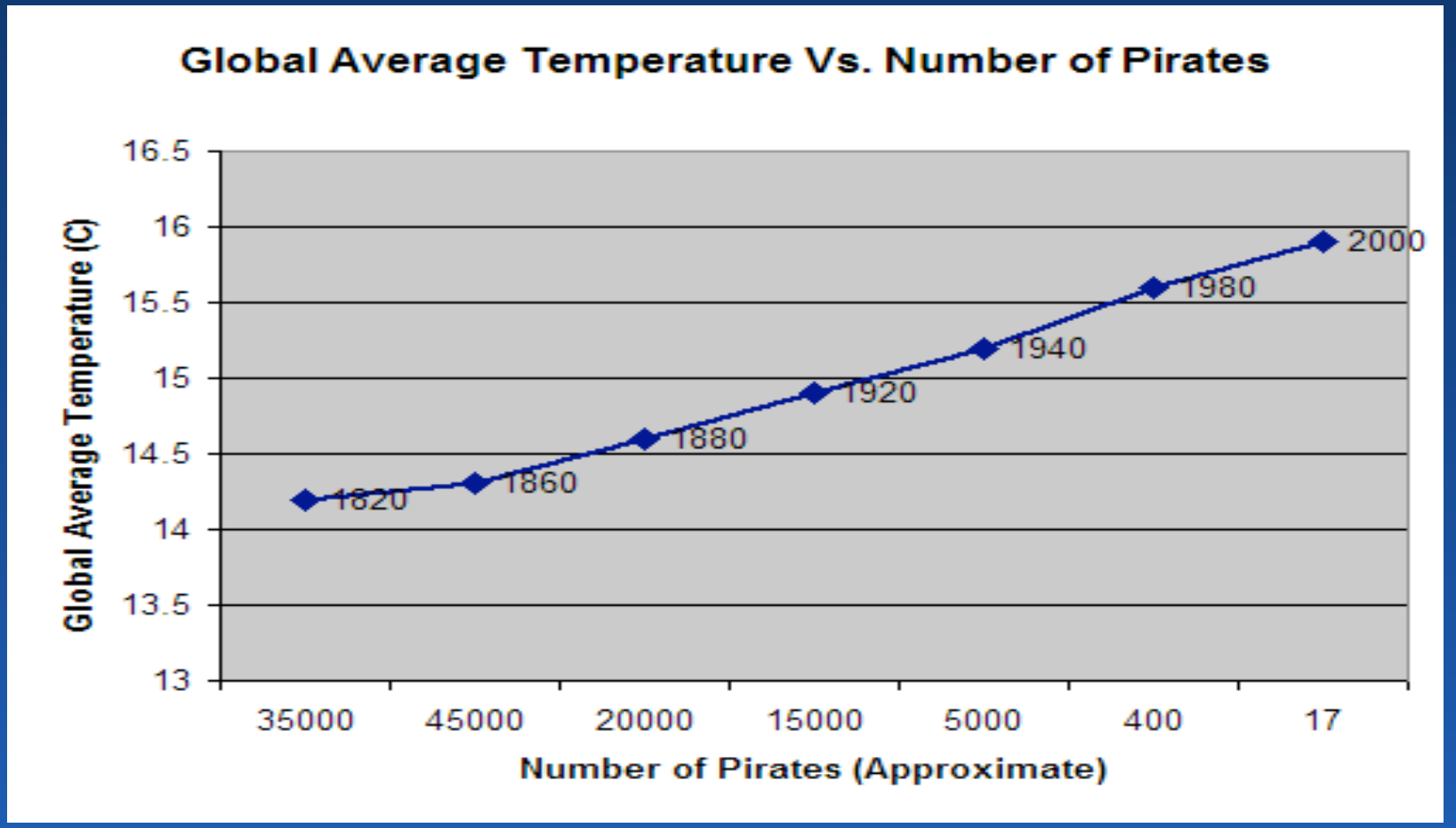
- R values also exist, but that's correlation (linear regression)

Statistics Simplicity to Complexity



Idaho National Laboratory

Statistics: Correlations and Causality



INL Idaho National Laboratory

Next Time

- **Bill Domke will present an overview of the history of proliferation around the world**